Threats to Federated Learning: A Survey

Year: 2020 | Citations: 500 | Authors: Lingjuan Lyu, Han Yu, Qiang Yang

---

Abstract

With the emergence of data silos and popular privacy awareness, the traditional centralized approach of training artificial intelligence (AI) models is facing strong challenges. Federated learning (FL) has recently emerged as a promising solution under this new reality. Existing FL protocol design has been shown to exhibit vulnerabilities which can be exploited by adversaries both within and without the system to compromise data privacy. It is thus of paramount importance to make FL system designers to be aware of the implications of future FL algorithm design on privacy-preservation. Currently, there is no survey on this topic. In this paper, we bridge this important gap in FL literature. By providing a concise introduction to the concept of FL, and a unique taxonomy covering threat models and two major attacks on FL: 1) poisoning attacks and 2) inference attacks, this paper provides an accessible review of this important topic. We highlight the intuitions, key techniques as well as fundamental assumptions adopted by various attacks, and discuss promising future research directions towards more robust privacy preservation in FL.